Volume: 15 Issue: 06 Pages: 6204 – 6210 (2024) ISSN: 0975-0290

Enhancing Outlier Detection and Dimensionality Reduction in Machine Learning for Extreme Value Analysis

Dr. Virendra Kumar Tiwari
Dept. of Computer Applications, Lakshmi Narain College of Technology (MCA), Bhopal, INDIA
Email: virugama@gmail.com
Ashish Jain
Dept. of Computer Applications, Lakshmi Narain College of Technology (MCA), Bhopal, INDIA
Email: Jainashish1302@gmail.com
Rohit Singh
Dept. of Computer Applications, Lakshmi Narain College of Technology (MCA), Bhopal, INDIA
Email: deepakrohit@gmail.com
Priyanka Singh
Dept. of Computer Engineering, Lakshmi Narain College of Technology Excellence, Bhopal, INDIA
Email: priyankasirt03@gmail.com

--**ABSTRACT**--

An outlier is generally defined as an observation that significantly deviates from the rest of the considered compliance. In the realm of ultramodern machine literacy, where analysis of multidimensional datasets is current, improving data quality is imperative for economists aiming to gain robust results. Numerous machine learning algorithms are sensitive to both the range and distribution of trait values within the input data. The presence of outliers in the input data can distort and misguide the training process of machine literacy algorithms, leading to prolonged training times, less precise models, and eventually inferior issues. Prior to the construction of prophetic models using training data, outliers can introduce incorrect representations, thereby impacting interpretations of the collected data. Traditional outlier discovery methods frequently concentrate on banning the tails of distributions and overlooking the data generation processes specific to individual datasets. Colorful styles are useful for detecting different types of outliers in high-dimensional datasets from two distinct perspectives: relating the devious aspects of a data object and setting devious data objects within a dataset. N-dimensional data poses a significant challenge within the realm of machine literacy, contributing to the diversity of challenges faced in the field.

Keywords **- Outlier detection, Machine Learning, High-Dimensional Data, Intrinsic Dimension (ID), k-Nearest Neighbor (k-NN).**

Date of Submission: February 18, 2024 Date of Acceptance: April 02, 2024

I. INTRODUCTION

One of the main challenges in outlier detection within large datasets is the necessity for the development of fully automated and efficient processes [1]. While large datasets present numerous challenges in terms of data processing and analysis, such as the difficulty of thorough examination, they also offer the opportunity to leverage the richness of the dataset for outlier detection. Despite the potentially greater uncertainty in large datasets compared to smaller ones (except in cases like surveys, where larger datasets allow for more precise answers), the sheer size of the dataset prompts the development of methods that can optimally utilize the available data.

A significant body of research has focused on identifying outliers in datasets composed of vectors of numerical or categorical attributes, known as 'multi-dimensional outliers' [2-4]. However, the identification of multi-dimensional outliers assumes a "clean" dataset where large deviations from expected values in individual variables characterize the true behavior of those variables, rather than indicating outliers. Conversely, if there are outliers in individual values, they can distort the detection of multi-dimensional outliers. Moreover, in high-dimensional datasets, some outliers may not be detected by multi-dimensional approaches as they may be "hidden" within the dataset's granularity and may not significantly affect the multi-dimensional distance metrics

Volume: 15 Issue: 06 Pages: 6204 – 6210 (2024) ISSN: 0975-0290

used in the outlier detection process. Hence, addressing the issue of outliers in individual variables is crucial. Detecting outliers involves identifying objects that significantly deviate from the typical distribution of the data.

Figure: Outlier Analysis for different predicted points

Identifying outlier points within main data streams holds significant value across various research domains, including the analysis and scrutiny of network traffic data such as connection-oriented records, web logs, wireless sensor networks, and financial transactions. Consequently, there is considerable interest in proposing new techniques that address the limitations of existing methods in outlier detection.

Many machine learning algorithms are highly sensitive to the range and distribution of attribute values in the input data. When outliers are present in the input data, they can distort the training process of machine learning algorithms, leading to longer training iteration times, less accurate models, and ultimately poorer results in outlier detection. Even before predictive models are constructed using training datasets, outliers can create ambiguous representations and distort the interpretation of collected data.

Moreover, outliers have the potential to skew the summary distribution of attribute values in descriptive statistics such as mean and standard deviation, as well as in visualizations like histograms and scatter- plots, thereby compressing the main body of the data and affecting the overall understanding of the dataset. Therefore, developing robust outlier detection techniques is crucial for ensuring the accuracy and reliability of data analysis processes across various domains.

II. LITERATURE SURVEY

The common problem of identifying outliers has been tackled through various approaches, which can broadly be categorized as global versus local outlier models. Global models make binary decisions about whether a given data object is an outlier or not. On the other hand, local outlier models assign a degree of outlines to each object, often represented by an "outlier factor" that quantifies how much of an outlier an object is. In many applications where ranking outliers in the database and retrieving the top-n outliers is desirable, a local outlier approach is preferred.

Another classification of outlier approaches distinguishes between supervised and unsupervised methods. In supervised approaches, the model is trained on a set of labeled data observations where the status of being an outlier or not is known. This requires learning the differences between various categories of unusual observations. Supervised methods are comprehensive but can present challenges, particularly when dealing with unbalanced classification tasks, such as finding outlier points.

In contrast, finding outlier points is often treated as an unsupervised problem when there is insufficient prior information for supervised learning methods. However, statistical methods for identifying outlier points rely on assumed distributions of objects. Various tests are conducted and optimized for each data distribution based on specific constraints, such as the expected number of outliers and the space where outliers are expected to occur. However, conventional approaches rely heavily on the assumption of a precise distribution, which may not always hold true.

Moreover, most analyses are univariate, examining a single attribute to identify outlier points. While some approaches combine predefined models and supervised learning methods, they still require knowledge of the distribution of objects beforehand. For instance, some methods assume the data consists of k Gaussian distributions and mean and standard deviations are calculated in a data-driven manner. However, these approaches may lack robustness as mean and standard deviation calculations are sensitive to outliers.

In [11], authors propose finding distances from the k-nearest neighbors and categorizing data objects based on their distances to their kth nearest neighbor. However, this approach only addresses time complexity concerns for high-dimensional data without focusing on the intrinsic difficulties of such data. Similarly, [10] introduces an approximation based on suggestion points.

In this paper [12], the author aims to develop an improved learning method for identifying outliers among normal observations. The core idea of this learning method is to utilize the local neighborhood information of an observation to determine whether it is an outlier or not. To precisely capture the neighborhood information, the concept of Local Projection Space (LPS) is introduced to compute the anomalous degree of a given observation. Formally, LPS aligns with the concept of nuclear norm and can be obtained through the process of low-rank matrix approximation.

Unlike other distance-based and density-based detection methods, the proposed method, known as LPOD (Local Projection Outlier Detection), demonstrates robustness to the parameter k of k-NN (k-Nearest Neighbors) embedded within LPOD. Through the application of this method, the authors effectively evaluate various outlier datasets. Experimental results indicate that LPS excels at ranking the most suitable candidates for individual outliers, and the performance of LPOD exhibits competency across various characteristics.
Additionally, the author introduces a new outlier scoring method, termed the Intrinsic Dimension Outlier Score (IDOS), to distinguish between inliers and outliers in the vicinity of a test position. The continuous intrinsic dimension (ID), which is shown to be equivalent to the discriminative power of similarity functions, plays a key role in estimating the local outliers. Observations with an increase in the estimated value of their continuous intrinsic dimension can be considered as local outliers. This concept extends Karger and Ruhl's expansion dimension to a statistical setting, where the sharing of distances to an uncertain point is modeled with a continuous random variable.

In comparison with the Local Outlier Factor (LOF), IDOS demonstrates a potential advantage in assessing local density within local clusters, making it easier to discriminate outliers about such clusters. Experimental analysis reveals that the precision of IDOS significantly improves the identification of outlier points based on scoring methods, especially when dealing with large and high-dimensional datasets. IDOS demonstrates its superiority in terms of both effectiveness and efficiency in handling such datasets.

As the dimensionality of data objects increases, it becomes increasingly challenging to identify data points that do not fit into any cluster, commonly referred to as outliers. This method of finding outlier points holds significant importance in real-life applications such as fraud detection, intrusion detection, and various other areas where data dimensions are expanding. In this context, the author proposes another method that involves dividing the original high-dimensional dataset into subspace clusters using a subspace clustering method. The goal is to improve the k-means algorithm by establishing an outlier cluster, which is then merged with other clusters based on a compromise task. Various outlier clusters that cannot be combined with any other subspace cluster are identified to determine the final outlier cluster.

In their study [14], the author investigates various research covering many concepts of high-dimensional data mining and information retrieval, focusing on finding outlier points in multidimensional data, ensemble subspace clustering, spam detection, and improving the k-means algorithm based on association rules. Since these types of data are crucial for information systems, these concepts can be utilized to enhance data mining and machine learning methods. These approaches are instrumental in designing robust applications for information retrieval.

One potential application highlighted is Spam Outlier Detection using Ensemble subspace clustering, where spam outliers in the analysis dataset of e-commerce platforms can be identified. Subspace clustering is performed, followed by outlier detection, and then the results are ensemble with other subspaces to achieve greater accuracy. By incorporating spam detection logic into the process, concerns regarding fraudulent reviews can be alleviated. Any clusters identified as outlier clusters from high-dimensional datasets can be highlighted or subject to appropriate actions.

Another proposed approach involves implementing elimination logic in datasets so that when outliers are detected initially, any incoming data belonging to the same dimension set will be rejected from being added to the database. This proactive measure helps maintain data integrity during analysis.

In this paper [15], the authors propose a hybrid semi-supervised anomaly detection model for high-dimensional data. The proposed detection model consists of two components: a deep autoencoder (DAE) and an ensemble k-nearest neighbor graph-based anomaly detector (K-NNG).
The deep autoencoder (DAE) leverages the advantages of nonlinear mapping methods. Initially, it is trained in an unsupervised mode to capture essential features of data objects and transform them into a high-dimensional space.

In this method, the training dataset is densely represented in the compact feature dimensional data space. This representation is then used with various nonparametric KNN-based anomaly detection methods, utilizing only a part of the real-life dataset rather than the entire specific training set. This approach significantly reduces computational costs.
Experimental results and statistical significance analysis demonstrate the effectiveness of the proposed method on several real-life datasets. The performance evaluation confirms that the hybrid model improves anomaly detection accuracy while also reducing computational complexity compared to standalone algorithms.

This proposal [16] is based on an extension of the mathematical framework upon which the basic theory of detection of outliers, founded on Rough Set Theory, has been constructed. From this starting point, current problems are analyzed; a detection method is proposed, along with a computational algorithm that allows the performance of outlier detection tasks with an almost-linear complexity. To illustrate its viability, the results of the application of the outlier-detection algorithm to the concrete example of a large database are presented.

Sewwandi et al. [17] proposed a novel approach for feature selection and classification using neighborhood rough set theory and k-nearest neighbor (k-NN) algorithms. Their method focuses on class-specific feature selection, aiming to improve classification accuracy by selecting the most relevant features for each class. By incorporating neighborhood rough set theory, which considers the local information of data points, their approach effectively handles complex data distributions and achieves promising results in various classification tasks.

In another study, Virendra Kumar Tiwari and Priyanka Singh [18] addressed the classification of motor imagery in EEG signals by employing feature optimization techniques and machine learning algorithms. Their work demonstrates the importance of feature selection and optimization in improving the performance of EEG-based classification systems. By identifying discriminative features related to motor imagery tasks and leveraging machine learning algorithms, they achieved accurate classification results, contributing to the advancement of brain-computer interface (BCI) systems.

Shukla et al. [19] conducted a comparative study of methods for internet traffic sharing in computer networks. Their work focuses on optimizing internet traffic distribution to enhance network performance and resource utilization. By comparing different sharing methods, including load balancing and traffic shaping techniques, they provide valuable insights into the trade-offs between efficiency, fairness, and scalability in network traffic management.

Thakur et al. [20] investigated internet traffic distribution management under cybercrime marketing plans. Their study sheds light on the challenges posed by malicious activities on the internet, such as distributed denial-of-service (DDoS) attacks and botnet-driven traffic. By analyzing various strategies for mitigating cyber threats and optimizing traffic distribution, they contribute to the development of effective cybersecurity measures and network defense mechanisms.

In a different context, Meng et al. [21] proposed a hybrid dimensionality reduction method for outlier detection in high-dimensional data. Their approach combines multiple dimensionality reduction techniques, such as principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE), to preprocess data before outlier detection. By reducing the dimensionality of high-dimensional datasets while preserving important information, their method improves the efficiency and effectiveness of outlier detection algorithms, facilitating anomaly detection tasks in diverse applications.

III. PROBLEM STATEMENT

The problem statement regarding outlier detection in streaming data and the associated challenges:

i. **Outlier Detection in Streaming Data**: Outliers are data points that significantly differ from most of the dataset. In the context of streaming data, which involves continuous and potentially infinite data streams, the challenge lies in identifying these outliers in real-time as new data points arrive. This requires algorithms capable of efficiently analyzing the incoming data stream and identifying deviations from the expected pattern.

ii. **Computational Complexity and Feature Reduction**: As streaming data often involves a high volume of incoming data points, the computational cost of analyzing and processing this data can be significant. Moreover, many outlier detection techniques rely on feature reduction algorithms to handle the high dimensionality of the data efficiently. However, these feature reduction techniques can themselves introduce computational overhead, especially when dealing with large datasets.

iii. **Efficient Feature Reduction Techniques**: To address the computational challenges associated with feature reduction, there is a need to develop techniques that can effectively reduce the dimensionality of the data while minimizing computational complexity. These techniques should be capable of handling the large volumes of incoming data inherent in streaming data scenarios without sacrificing accuracy or performance.

iv. **Predictive and Direct Outlier Detection Techniques**: Outlier detection algorithms can be categorized into predictive and direct techniques. Predictive techniques utilize labeled data from training sets to build models that can predict outliers based on patterns observed in the training data. These models are then applied to classify original data objects as outliers or non-outliers. Direct techniques, on the other hand, directly analyze the data to identify outliers without the need for labeled training data.

Outlier detection from streaming data involves identifying items, such as objects or points, that deviate significantly from the norm within the entire data set or a defined window of the data stream. Another challenge lies in the computational cost associated with feature reduction algorithms, especially considering the large volumes of incoming data. Therefore, there is a need to develop computationally efficient feature reduction techniques that can handle the influx of data effectively.

Various algorithms for outlier detection are based on statistical modelling techniques, which can be predictive or direct in nature. Predictive techniques utilize labelled data from training sets to build a model for outlier detection, which is then applied to classify original data objects. These techniques aim to predict outliers based on patterns observed in the training data, allowing for the identification of abnormal or irregular data points.

IV. MACHINE LEARNING PRINCIPLES

Machine learning (ML) is the process of training a computer (machine) to acquire knowledge from data, akin to gaining experience, with the goal of constructing a framework for generating reliable, self-improving, accurate predictions, and identifying patterns from high-dimensional data in each task. It is particularly useful for handling sets of tasks that are too difficult, complex, or error-prone for humans to perform. Machine learning algorithms can generally be categorized into the following three groups [5]:

1. **Supervised learning** Supervised learning encompasses algorithms that learn from labeled data. Each instance, denoted as Xj, consists of input features, and yj represents the corresponding target. The algorithm automatically learns a function f that maps inputs to their desired targets, expressed as f(Xj)=yj. When the target values belong to a set of categories, the problem is termed a classification problem. Alternatively, when targets have continuous values, it becomes a regression problem.
2. **Unsupervised learning** Unsupervised learning refers to algorithms that learn from unlabeled inputs Xj. The objective is to automatically identify and understand patterns and structures concealed within Xj. One crucial task in unsupervised learning is clustering, where similar instances are grouped together in the same cluster.
3. **Semi-Supervised learning** Semi-supervised learning lies between supervised and unsupervised learning. These algorithms learn from both labeled and unlabeled instances. Depending on the task, methods from unsupervised and supervised learning are adapted for use in semi-supervised learning. Thus, semi-supervised learning not only learns from labeled instances and their associations with desired targets but also potentially exploits hidden correlations between labeled and unlabeled instances if such correlations exist.

V. TYPES OF OUTLIER DETECTION TECHNIQUES

Outliers can be classified into two main varieties: univariate and multivariate. Uni-variate outliers are identified by examining the distribution of values within a single feature space. On the other hand, multivariate outliers are identified within an n-dimensional space, where n represents the number of features.

Detecting outlier points within distributions in n-dimensional spaces can be extremely challenging for the human brain, underscoring the need to train models to perform this task efficiently.

Moreover, outliers can manifest in different types depending on the context: point outliers, contextual outliers, or collective outliers. Point outliers are individual data points that deviate significantly from the rest of the distribution. Contextual outliers may arise as noise in data, such as punctuation symbols in text analysis or background noise in speech recognition. Collective outliers represent subsets of uniqueness within the data, indicating novel occurrences.

Additionally, Aguinis et al. (2013) [6] propose a classification of outliers into three categories: error outliers, interesting outliers, and influential outliers. Here, we introduce two modifications to their classification.

- **Influential outliers,** as defined by Aguinis et al. (2013), are outliers that exert significant influence either on the model fit (model fit outliers) or on parameter estimation (prediction outliers). Model fit outliers become apparent when using statistical methods based on maximum likelihood (and its variants). Prediction outliers, on the other hand, become noticeable when utilizing more common methods like least squares (as seen in linear regression).
- **Error outliers** are observations that deviate from other data points due to inaccuracies, as stated by Aguinis et al. (2013). These inaccuracies can stem from dimension errors or encoding errors. Both error and interesting outliers are influenced by moderators. The moderator of an error outlier is typically identified as lacking theoretical relevance and arises from an error (e.g., a coding error). On the other hand, an interesting outlier is influenced by a moderator that may or may not be identified and could potentially be of theoretical interest.
- **Interesting outliers** are not explicitly errors but may be influenced by potentially intriguing moderators. These moderators may or may not be of theoretical

Volume: 15 Issue: 06 Pages: 6204 – 6210 (2024) ISSN: 0975-0290

interest and could remain undisclosed. Therefore, it may be more appropriate to refer to them as potentially interesting outliers. Random outliers are values that occur purely by chance, such as a completely fair coin landing on "heads" 100 times in 100 throws. Random outliers are inherently suspicious but still possible. Considering typical cutoffs for outlier detection, one would expect no more than 0.27% of random outliers.

VI. PROPOSED SOLUTIONS:

Outlier detection in streaming data involves identifying items deviating significantly from the dataset norm or a defined window of the data stream. A key challenge is the computational cost associated with feature reduction algorithms, especially given the high data volume. Thus, there's a necessity for computationally efficient feature reduction techniques capable of managing data influx effectively.

Algorithms for outlier detection commonly rely on statistical modeling, categorized as predictive or direct. Predictive techniques utilize labeled data to build models for outlier detection, facilitating the identification of abnormal data points based on observed patterns.

VII. CONCLUSION

Outliers are extreme values that deviate significantly from other observations in a dataset; they may indicate errors in measurement, experimental anomalies, or novel phenomena. Detecting outlier points is a crucial aspect of machine learning tasks with numerous critical applications, such as medical diagnosis, fraud detection, and intrusion detection. However, in real-life applications with large datasets, outlier detection faces various challenges. To address these challenges, efficient dimensionality reduction techniques are employed to reduce the dimensionality of the data while preserving valuable information.

Dimensionality reduction aims to transform data into a lower-dimensional space while minimizing information loss. Nonlinear transformations offer greater flexibility in aligning data along fewer dimensions but may also disrupt the data's structure. t-SNE (t-Distributed Stochastic Neighbor Embedding) is a technique used to visualize high-dimensional data. Unlike many other dimensionality reduction methods that focus on preserving global structure, t-SNE aims to group local data points closer together, making it more aligned with human reasoning. This algorithm computes pairwise conditional probabilities and minimizes the difference between these probabilities in higher and lower dimensions.

Given the computational complexity involved in t-SNE calculations, it offers a more efficient approach compared to other dimensionality reduction techniques. Extracting valuable information from data is increasingly important, particularly in finding patterns within large datasets. Machine learning and statistical techniques play a crucial role in improving performance across various domains within machine learning.

REFERENCES

[1]. Maciá-Pérez F., Berna-Martinez J., Fernández Oliva A., Ortega, M. Abreu, 2015. Algorithm for the detection of outliers based on the theory of rough sets. Decision Support Systems, 75, pp. 63-75.

[2]. Otey, M., Ghoting, A., Parthasarathy, S., 2006. Fast distributed outlier detection in mixed-attribute data sets. Data Mining and Knowledge Discovery 12, 203-228.

[3]. Koufakou, A., Georgiopoulos, M., 2010. A fast outlier detection strategy for distributed high- dimensional data sets with mixed attributes. Data Mining and Knowledge Discovery, 20, 259-289.

[4]. Kutsuma, T., Yamamoto, A. 2017. Outlier detection using binary decision diagrams. Data Mining and Knowledge Discovery, 31, 548-572.

[5]. Murphy, K. P.: Machine learning: a probabilistic perspective. 2012.

[6]. Aguinis, H., Gottfredson, R.-K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, 16(2), 270–301. DOI: https://doi.org/10.1177/1094428112470848.

[7]. C. Zhu, H. Kitagawa, and C. Faloutsos. Example-based robust outlier detection in high dimensional datasets. In Proc. ICDM, 2005.

[8]. G. Williams, K. Yamanishi, and J. Takeuchi. Online unsupervised outlier detection using finite mixtures with discounting learning algorithms. In Proc. KDD, 2000.

[9]. K. Yamanishi and J. Takeuchi. Discovering outlier filtering rules from unlabeled data: combining a supervised learner with an unsupervised learner. In Proc. KDD, 2001.

[10]. Y. Pei, O. Zaıane, and Y. Gao. An efficient reference-based approach to outlier detection in large datasets. InProc. ICDM, 2006.

[11]. S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. InProc. SIGMOD, 2000.

[12]. Huawen Liu, Member, IEEE, Xuelong Li, Fellow, IEEE, Jiuyong Li, Member, IEEE, and Shichao Zhang, Senior Member, IEEE "Efficient Outlier Detection for High-Dimensional Data" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, 2017.

[13]. Jonathan von Brunken, Michael E. Houle, and Arthur Zimek, "Intrinsic Dimensional Outlier Detection in High-Dimensional Data" NII-2015 -003E, Mar. 2015.

[14]. Suresh S. Kapare, Bharat A. Tidke, "Spam Outlier Detection in High Dimensional Data: Ensemble Subspace Clustering Approach" IJCSIT) International Journal of

Computer Science and Information Technologies, Vol. 6 (3), 2015, 2326-2329.

[15]. Hongchao Song, Zhuqing Jiang, Aidong Men, and Bo Yang, “A Hybrid Semi-Supervised Anomaly Detection Model for High Dimensional Data” Comput Intell Neurosci. 2017.

[16]. M.A.N.D. Sewwandi, Yuefeng Li, Jinglan Zhang, “k-outlier removal based on contextual label information and cluster purity for continuous data classification” Expert Systems with Applications, Volume 237, Part C, 1 March 2024, 121347.

[17]. Sewwandi M., Li Y., Zhang J., “A class-specific feature selection and classification approach using neighborhood rough set and k-nearest neighbor theories”, Applied Soft Computing, 1568-4946, 143 (2023), p. 110366,

[18]. Dr. Virendra Kumar Tiwari, Priyanka Singh “Classification of Motor Imaginary in EEG using feature Optimization and Machine Learning” International Journal of Advanced Networking and Applications (IJANA), India, vol. 15, issue 02, pp. 5887– 5891, August 2023. DOI: 10.35444/IJANA. 2023.15208.

[19]. Shukla D., Tiwari Virendra, Thakur S. and Tiwari M. “A comparison of methods for Internet traffic sharing in computer network”, International Journal of Advanced Networking and Applications (IJANA), India, vol. 1, issue 3, pp. 164-169(06), 2009.

[20]. Thakur Sanjay, Jain Saurabh, Tiwari Virendra, and Shukla D., (June 2014): Internet Traffic Distribution Management under Cyber Crime Marketing Plans, International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), India, vol. 4, issue 6, pp. 185-193(09), June 2014.

[21]. Guanglei Meng, Biao Wang, Yanming Wu, Mingzhe Zhou & Tiankuo Meng “A hybrid dimensionality reduction method for outlier detection in high-dimensional data” springer link, International Journal of Machine Learning and Cybernetics, Volume 14, pages 3705–3718, (2023)

Biographies and Photographs

Dr. Virendra Kumar Tiwari earned his B.Sc., M.A. (Economics), MCA, and a Ph.D. in Computer Science from Dr. Hari Singh Gour University, Sagar, Madhya Pradesh. He is a prolific researcher, having authored numerous papers published in esteemed National and International Journals. Currently, Dr. Virendra Kumar Tiwari holds the position of Professor in the Department of Computer Applications within the Faculty of Computer Science & Engineering at Lakshmi Narain College of Technology in Bhopal, Madhya Pradesh. Dr. Virendra Kumar Tiwari has played a significant role in organizing various conferences and seminars. Additionally, he has published three books covering diverse subjects such as "Research and Applications Towards Mathematics and Computer Science," "Advance Computer Network," and "Data Analytics." Furthermore, he holds patent and copyright. His areas of expertise encompass Computer Networks and Databases.

Professor **Ashish Jain** holds a master’s degree in computer applications from Rajeev Gandhi Vishwavidyalaya. He is the author of numerous publications in esteemed national and international journals. Currently, he serves as an Associate Professor in the Department of Computer Applications within the Faculty of Computer Science & Engineering at Lakshmi Narain College of Technology in Bhopal, Madhya Pradesh.

Professor **Rohit Singh** holds a master’s degree in computer applications from SATI Vidisha at Rajeev Gandhi Vishwavidyalaya. He has authored numerous publications in esteemed national and international journals. Currently, he serves as an Associate Professor in the Department of Computer Applications within the Faculty of Computer Science & Engineering at Lakshmi Narain College of Technology in Bhopal, Madhya Pradesh.

Professor **Priyanka Singh** earned her MCA and M. Tech. (CSE) degrees from Rajiv Gandhi Proudyogiki Vishwavidyalaya in Bhopal. Since 2022, she has been serving as a research scholar in the Department of Computer Science and Engineering at Lakshmi Narain College of Technology Excellence in Bhopal, Madhya Pradesh. Her research focuses on IoT, Machine Learning, and Deep Learning, employing various analysis techniques.